

競馬の数理モデル

2022年4月24日

目次

1	はじめに	2
2	単勝の確率	2
3	順序確率	3
4	ポートフォリオ	4
4.1	シンプルな定式化	4
4.2	少し複雑な定式化	8
5	機械学習	16
	参考文献	17

1 はじめに

競馬の数理モデルを紹介します。様々な数理モデルがありますが、本稿の第2節から第4節では非常にシンプルなモデルを取上げます。

2 単勝の確率

今 $n \geq 2$ 頭が出走するレースがあるとする。出走馬には $1, 2, \dots, n$ とそれぞれ番号が付けられている。馬 i が1着になる確率を π_i とおく。このとき、各馬 i の勝率 π_i がどの程度かが問題になる。レースは Parimutuel 方式で開催されるとする。Parimutuel 方式はギャンブルにおけるオッズを定める方式の1つである。Parimutuel 方式では、競馬場が馬券の総売上げ額から一定割合を抜き取り、残りの金額を的中馬券に配分し、オッズを定める。勝率 π_i を推定する最もシンプルなモデルとして

$$\pi_i = P_i \quad (i = 1, 2, \dots, n)$$

がある。ここで P_i は単勝馬券総売上げ額に対する単勝馬券 i の総売上げ額の割合を表す。 P_i は以下の式で計算できる。

$$P_i = \frac{1}{\sum_{j=1}^n \frac{1}{O_{j+1}}} \quad (i = 1, 2, \dots, n). \quad (1)$$

ここで O_i は馬 i の単勝オッズを表す。Lo and Bacon-Shone (2008b) では上記モデルの改良版として以下のモデルを紹介している。

$$\pi_i = \frac{P_i^\beta}{\sum_{j=1}^n P_j^\beta}, \quad (i = 1, 2, \dots, n).$$

ここで β はパラメータを表し、値はおおよそ 1.1 になる。 $\beta = 1$ のとき、 $\pi_i = P_i$ が成立し最もシンプルなモデルとなる。 β の詳細な値は Lo and Bacon-Shone (2008b) の Table 1 に記載されている。

続いてこの段落では Ali (2008) に基づき、 P_i と単勝オッズ O_i の関係式 (1) を導出する。まず Parimutuel 方式により、馬 i の単勝オッズ O_i は以下で定められている。

$$O_i = \frac{(1-t)W}{W_i} - 1 \quad (i = 1, 2, \dots, n). \quad (2)$$

ここで、 W は単勝馬券の総売上げ額を、 W_i は単勝馬券 i の総売上げ額を、 t は競馬場の取分の割合をそれぞれ表す。よって、

$$W = \sum_{i=1}^n W_i \quad (3)$$

が成立する。ここで $P_i = \frac{W_i}{W}$ を単勝オッズのみで表すことを考える。式 (2) より

$$\frac{1}{O_i + 1} = \frac{W_i}{(1-t)W} \quad (i = 1, 2, \dots, n)$$

なので、

$$\begin{aligned} (1-t) \sum_{i=1}^n \frac{1}{O_i + 1} &= (1-t) \sum_{i=1}^n \frac{W_i}{(1-t)W} \\ &= \frac{1}{W} \sum_{i=1}^n W_i \\ &= 1. \end{aligned} \quad \because \text{式 (3) より}$$

上式より、

$$1-t = \frac{1}{\sum_{i=1}^n \frac{1}{O_i + 1}} \quad (4)$$

が成立する。よって

$$\begin{aligned} P_i = \frac{W_i}{W} &= \frac{(1-t)}{O_i + 1} && \because \text{式 (2) より} \\ &= \frac{\frac{1}{O_i + 1}}{\sum_{j=1}^n \frac{1}{O_j + 1}}. && \because \text{式 (4) より} \end{aligned}$$

以上より、関係式 (1) の導出が終了した。

3 順序確率

この節では、レースにおける2着、3着がどのような状況になるのかを説明する。Harvilleモデル、ディスカウントモデルと呼ばれる2つの確率モデルを紹介する。馬 i が1着かつ馬 j が2着である確率を π_{ij} とおく。さらに馬 i が1着かつ馬 j が2着かつ馬 k が3着である確率を π_{ijk} とおく。Harvilleは次のシンプルなモデルを提案した。

$$\begin{aligned} \pi_{ij} &= \pi_i \frac{\pi_j}{1 - \pi_i}, \\ \pi_{ijk} &= \pi_i \frac{\pi_j}{1 - \pi_i} \frac{\pi_k}{1 - \pi_i - \pi_j}. \end{aligned}$$

そして, Lo and Bacon-Shone (2008a) は上記を包含する次のモデルを提案した.

$$\pi_{ijk} = \pi_i \frac{\pi_j^\lambda}{\sum_{s \neq i} \pi_s^\lambda} \frac{\pi_k^\tau}{\sum_{t \neq i, j} \pi_t^\tau}.$$

λ と τ はパラメータ. このモデルはディスカウントモデルと呼ばれている. $\lambda = \tau = 1$ のときディスカウントモデルは Harville のモデルになる. パラメータ λ, τ の一例として, $\lambda = 0.76, \tau = 0.62$ を挙げておく.

4 ポートフォリオ

この節ではギャンブラーがどの馬券をどれだけ購入すれば良いのかを問題にする. つまり, 馬券によるポートフォリオの構成を問題にする. 一般にポートフォリオは Kelly criterion に基づいて構成されることが多い. この節では 2 種類の定式化を行い, それぞれに対して Kelly criterion を導出する.

4.1 シンプルな定式化

この節では Tijms (2012) の 2.7 節にあるシンプルな Kelly criterion を紹介する. 今ギャンブルに参加することを考える. 勝ったときは払戻し率 u で賭金が払戻され, 負けたときは賭金すべてを失うとする. ここで $u > 1$ を仮定する. 勝つ確率を p , 負ける確率を $1 - p$ とする. そして, $0 < p < 1$ を仮定する. 所持金のうち賭金の割合を f とする. 所持金のうち賭金にせず残す金額の割合を w とする. このとき所持金をどのように配分し賭けるかが問題になる. この問題を以下の連続最適化問題に定式化する.

問題 1.

$$\begin{aligned} & \text{maximize} && \mu(f, w) := p \log(uf + w) + (1 - p) \log w \\ & \text{subject to} && f + w = 1, \\ & && f \geq 0, \\ & && w \geq 0 \end{aligned} \tag{5}$$

この問題は, $(f, w) = (f^*, w^*)$ のとき最適解を持つものとする. まず

$$pu > 1 \tag{6}$$

の場合を考える. 以下では Körner (2008) の Exercise 2.7.2 を参考にして補題 1 及び定理 1 の証明 1 を導出する.

補題 1. 以下のすべてが成立する.

1.

$$w^* > 0$$

2.

$$f^* > 0$$

3.

$$p \frac{u-1}{uf^* + w^*} = (1-p) \frac{1}{w^*} \quad (7)$$

Proof. 1. $w^* > 0$ は自明. もしも $w^* = 0$ ならば、 (f^*, w^*) の最適性に矛盾してしまう.

2.

$$h(x) := \mu(f^* + x, w^* - x)$$

とおく. すると

$$h(x) = p \log((u-1)x + uf^* + w^*) + (1-p) \log(-x + w^*).$$

上式を微分すると

$$h'(x) = p \frac{u-1}{(u-1)x + uf^* + w^*} + (1-p) \frac{-1}{-x + w^*}.$$

したがって

$$h'(0) = p \frac{u-1}{uf^* + w^*} - (1-p) \frac{1}{w^*}.$$

今 $f^* = 0$ を仮定して矛盾を導く. $h'(0) \leq 0$ なので

$$p \frac{u-1}{w^*} \leq (1-p) \frac{1}{w^*}.$$

上式より $pu \leq 1$ となるが、これは式 (6) に矛盾する. よって $f^* > 0$ が成立する.

3. $f^* > 0$ かつ $w^* > 0$ なので、 $h'(0) = 0$ となり

$$p \frac{u-1}{uf^* + w^*} = (1-p) \frac{1}{w^*}$$

が成立する. □

定理 1. $pu > 1$ のとき、問題 1 は以下の最適解 (f^*, w^*) を持つ.

$$f^* = \frac{pu-1}{u-1} = p - \frac{w^*}{u}$$

$$w^* = \frac{1-p}{1-\frac{1}{u}}$$

(証明 1). 式 (7) より

$$\begin{aligned} 0 &= w^*p(u-1) - (1-p)(uf^* + w^*) \\ &= \{p(u-1) - (1-p)\}w^* - (1-p)uf^* \\ &= (pu-1)w^* - (1-p)uf^*. \end{aligned}$$

上式に $w^* = 1 - f^*$ を代入すると、

$$\begin{aligned} 0 &= (pu-1)(1-f^*) - (1-p)uf^* \\ &= \{-(pu-1) - (1-p)u\}f^* + (pu-1) \\ &= (1-u)f^* + (pu-1). \end{aligned}$$

よって

$$f^* = \frac{pu-1}{u-1}$$

が成立する. そして、

$$\begin{aligned} w^* &= 1 - f^* \\ &= 1 - \frac{pu-1}{u-1} \\ &= \frac{u-pu}{u-1} \\ &= \frac{1-p}{1-\frac{1}{u}}. \end{aligned}$$

つまり

$$w^* = \frac{1-p}{1-\frac{1}{u}}$$

が成立する. さらに、

$$\begin{aligned} f^* &= \frac{pu-1}{u-1} \\ &= p + \frac{p-1}{u-1} \\ &= p - \frac{1}{u} \frac{1-p}{1-\frac{1}{u}} \\ &= p - \frac{w^*}{u}. \end{aligned}$$

つまり

$$f^* = p - \frac{w^*}{u}$$

が成立する. □

次に、Körner (2008) の Lemma 2.6.4 に基づいて、定理 1 に別証明を与える。

(定理 1 の証明 2). 問題 1 の最適解を求めるには、 $g(f) := \mu(f, 1-f)$ とおき、 $0 \leq f < 1$ における最大値を求めれば良い。

$$g(f) = p \log((u-1)f+1) + (1-p) \log(1-f)$$

を微分すると、

$$\begin{aligned} g'(f) &= p \frac{u-1}{(u-1)f+1} + (1-p) \frac{-1}{1-f} \\ &= \frac{p(u-1)(1-f) - (1-p)\{(u-1)f+1\}}{\{(u-1)f+1\}(1-f)}. \end{aligned}$$

ここで、

$$\begin{aligned} (\text{上式の分子}) &= \{-p - (1-p)\}(u-1)f + p(u-1) - (1-p) \\ &= -(u-1)f + (pu-1) \\ &= -(u-1) \left(f - \frac{pu-1}{u-1} \right). \end{aligned}$$

つまり、

$$g'(f) = \frac{-(u-1) \left(f - \frac{pu-1}{u-1} \right)}{\{(u-1)f+1\}(1-f)}. \quad (8)$$

ここで $pu > 1$ より、 $\frac{pu-1}{u-1} > 0$. さらに

$$\frac{pu-1}{u-1} - 1 = \frac{pu-1-(u-1)}{u-1} = \frac{-u(1-p)}{u-1} < 0$$

より、 $\frac{pu-1}{u-1} < 1$. したがって $g(f)$, $0 \leq f < 1$ は $f = \frac{pu-1}{u-1}$ のとき最大値をとる. つまり問題 1 は

$$f^* = \frac{pu-1}{u-1}$$

において最適解を持つ. そして、

$$w^* = \frac{1-p}{1-\frac{1}{u}}, \quad f^* = p - \frac{w^*}{u}$$

の証明は証明 1 と全く同じなので省略する. □

つぎに $pu \leq 1$ の場合を考える.

定理 2. $pu \leq 1$ のとき、問題 1 は最適解 $(f^*, w^*) = (0, 1)$ をもつ.

証明 1. (f, w) を任意の実行可能解とする. \log は上に凸な関数なので、Jensen の不等式を $\mu(f, w)$ へ適用すると、

$$\begin{aligned}\mu(f, w) &= p \log(uf + w) + (1 - p) \log w \\ &\leq \log \{p(uf + w) + (1 - p)w\} \\ &= \log(puf + w).\end{aligned}$$

さらに、

$$\begin{aligned}\log(puf + w) &\leq \log(f + w) && \because pu \leq 1 \text{ より} \\ &\leq \log 1 && \because \text{式 (5) より} \\ &= 0\end{aligned}$$

つまり任意の実行可能解 (f, w) に対して、 $\mu(f, w) \leq 0$ となる. $\mu(0, 1) = 0$ なので、問題 1 は $(f^*, w^*) = (0, 1)$ において最適解を持つ. \square

次に、Körner (2008) の Lemma 2.6.4 に基づいて、定理 2 に別証明を与える.

定理 2 の証明 2. $pu \leq 1$ のとき、 $0 < f < 1$ を満たす f に対して式 (8) は $g'(f) < 0$ となる. したがって $g(f)$, $0 \leq f < 1$ は $f = 0$ のとき最大値をとる. つまり、問題 1 は $(f^*, w^*) = (0, 1)$ において最適解を持つ. \square

4.2 少し複雑な定式化

この節では Ethier (2010) の Example 10.2.3 及び Körner (2008) の Exercise 2.7.2 を整理する. 4.1 節より少し複雑な定式化をおこない、Kelly の criterion を導出する. 今単勝馬券のみを購入するギャンブルを考える. レースでは $n \geq 2$ 頭が出走し、同着はないものとする. 馬には $1, 2, \dots, n$ と番号が付けられている. 馬 i が 1 着になる確率を p_i とする. ここで、 $p_1 > 0, p_2 > 0, \dots, p_n > 0$ を仮定する. そして $\sum_{i=1}^n p_i = 1$. 馬 i が 1 着になったときの払戻し率を $u_i > 1$ とおく. ここで

$$\sum_{i=1}^n \frac{1}{u_i} > 1 \tag{9}$$

を仮定する. 一般性を失わず、

$$u_1 p_1 \geq u_2 p_2 \geq \dots \geq u_n p_n \tag{10}$$

が満たされるように馬番号を並べ替える. 所持金のうち馬券 i を購入する割合を f_i とする. 所持金のうち現金として残す割合を w とする. ここで、 n 種類の単勝馬券をどのように購入し所持金を分配するかが問題になる. この問題を以下の連続最適化問題に定式化する.

問題 2.

$$\begin{aligned}
& \text{maximize} \quad \mu(\mathbf{f}, w) := \sum_{i=1}^n p_i \log(u_i f_i + w) \\
& \text{subject to} \quad \sum_{i=1}^n f_i + w = 1, \\
& \quad \quad \quad f_i \geq 0, \quad i = 1, 2, \dots, n, \\
& \quad \quad \quad w \geq 0
\end{aligned} \tag{11}$$

この問題は、 $(\mathbf{f}, w) = (\mathbf{f}^*, w^*)$ のとき最適解を持つものとする. まず,

$$u_1 p_1 > 1 \tag{12}$$

の場合を考える.

補題 2. $1 \leq i < j \leq n$ とする.

1. $f_i^* \neq 0$ かつ $f_j^* \neq 0$ のとき

$$\frac{p_i u_i}{u_i f_i^* + w^*} = \frac{p_j u_j}{u_j f_j^* + w^*}. \tag{13}$$

2. $f_i^* = 0$ のとき

$$f_j^* = 0. \tag{14}$$

3. $f_j^* = 0$ のとき

$$\frac{p_i u_i}{u_i f_i^* + w^*} \geq \frac{p_j u_j}{w^*}. \tag{15}$$

Proof.

$$g(x) := \mu(f_1^*, \dots, f_{i-1}^*, f_i^* + x, f_{i+1}^*, \dots, f_{j-1}^*, f_j^* - x, f_{j+1}^*, \dots, f_n^*, w^*)$$

とおく. すると

$$g(x) = p_i \log(u_i (f_i^* + x) + w^*) + p_j \log(u_j (f_j^* - x) + w^*) + \sum_{\substack{k=1 \\ k \neq i, j}}^n p_k \log(u_k f_k^* + w^*).$$

上式を微分すると、

$$\begin{aligned} g'(x) &= \frac{d}{dx} (p_i \log(u_i x + u_i f_i^* + w^*) + p_j \log(-u_j x + u_j f_j^* + w^*)) \\ &= \frac{p_i u_i}{u_i x + u_i f_i^* + w^*} - \frac{p_j u_j}{-u_j x + u_j f_j^* + w^*}. \end{aligned}$$

したがって

$$g'(0) = \frac{p_i u_i}{u_i f_i^* + w^*} - \frac{p_j u_j}{u_j f_j^* + w^*}.$$

1. $f_i^* \neq 0$ かつ $f_j^* \neq 0$ のとき. $g'(0) = 0$ なので

$$\frac{p_i u_i}{u_i f_i^* + w^*} = \frac{p_j u_j}{u_j f_j^* + w^*}.$$

2. $f_i^* = 0$ のとき. $g'(0) \leq 0$ なので

$$\frac{p_i u_i}{w^*} \leq \frac{p_j u_j}{u_j f_j^* + w^*}.$$

今 $f_j^* > 0$ を仮定すると、

$$\frac{p_i u_i}{w^*} < \frac{p_j u_j}{0 + w^*}$$

より、 $p_i u_i < p_j u_j$ となるが、これは式 (10) に矛盾する。よって $f_j^* = 0$ 。

3. $f_j^* = 0$ のとき. $g'(0) \geq 0$ なので

$$\frac{p_i u_i}{u_i f_i^* + w^*} \geq \frac{p_j u_j}{w^*}.$$

□

補題 3. 以下のすべてが成立する。

1.

$$f_1^* > 0 \tag{16}$$

2.

$$w^* > 0 \tag{17}$$

3.

$$\frac{p_1 u_1}{u_1 f_1^* + w^*} = \sum_{i=1}^n \frac{p_i}{u_i f_i^* + w^*} \tag{18}$$

4.

$$f_n^* = 0 \tag{19}$$

Proof.

$$h(x) := \mu(f_1^* + x, f_2^*, \dots, f_n^*, w^* - x)$$

とおく. すると

$$h(x) = p_1 \log((u_1 - 1)x + u_1 f_1^* + w^*) + \sum_{i=2}^n p_i \log(-x + u_i f_i^* + w^*).$$

上式を微分すると、

$$h'(x) = p_1 \frac{u_1 - 1}{(u_1 - 1)x + u_1 f_1^* + w^*} + \sum_{i=2}^n p_i \frac{-1}{-x + u_i f_i^* + w^*}$$

したがって、

$$\begin{aligned} h'(0) &= p_1 \frac{u_1 - 1}{u_1 f_1^* + w^*} - \sum_{i=2}^n p_i \frac{1}{u_i f_i^* + w^*} \\ &= \frac{p_1 u_1}{u_1 f_1^* + w^*} - \sum_{i=1}^n \frac{p_i}{u_i f_i^* + w^*}. \end{aligned}$$

1. $f_1^* = 0$ を仮定して矛盾を導く. 式 (14) より $f_2^* = 0, f_3^* = 0, \dots, f_n^* = 0$ となり $w^* = 1$ となる. 今 $h'(0) \leq 0$ なので、

$$p_1 u_1 \leq \sum_{i=1}^n p_i.$$

つまり $p_1 u_1 \leq 1$ となるが、これは式 (12) に矛盾する. よって $f_1^* > 0$ が成立する.

2. $w^* = 0$ を仮定して矛盾を導く. $w^* = 0$ なので最適性より $f_1^* > 0, f_2^* > 0, \dots, f_n^* > 0$ となる. すると式 (13) より

$$\frac{p_1}{f_1^*} = \frac{p_2}{f_2^*} = \dots = \frac{p_n}{f_n^*}.$$

上式の値を α とおくと、 $p_1 = \alpha f_1^*, p_2 = \alpha f_2^*, \dots, p_n = \alpha f_n^*$. これらを合計すると、

$$\begin{aligned} 1 &= \sum_{i=1}^n p_i \\ &= \sum_{i=1}^n \alpha f_i^* \\ &= \alpha \sum_{i=1}^n f_i^* \\ &= \alpha. \end{aligned} \quad \because \text{式 (11) より}$$

よって $\alpha = 1$ なので $f_1^* = p_1, f_2^* = p_2, \dots, f_n^* = p_n$ が成立する. 今 $h'(0) \geq 0$ なので

$$1 \geq \sum_{i=1}^n \frac{1}{u_i}$$

となるが、これは式 (9) に矛盾する. よって $w^* > 0$ が成立する.

3. $f_1^* > 0$ かつ $w > 0$ なので、 $h'(0) = 0$ となり、

$$\frac{p_1 u_1}{u_1 f_1^* + w^*} = \sum_{i=1}^n \frac{p_i}{u_i f_i^* + w^*}$$

が成立する.

4. $f_n^* > 0$ を仮定して矛盾を導く. $f_n^* > 0$ なので式 (14) より、 $f_1^* > 0, f_2^* > 0, \dots, f_n^* > 0$. 式 (13) の値を β とおけば、式 (18) は

$$\beta = \sum_{i=1}^n \frac{\beta}{u_i} = \beta \sum_{i=1}^n \frac{1}{u_i}$$

となる. $\beta > 0$ より上式は $1 = \sum_{i=1}^n \frac{1}{u_i}$ となるが、これは式 (9) に矛盾する. よって $f_n^* = 0$ が成立する. \square

$1 \leq k \leq n$ を満たす k に対して、

$$q_k := \sum_{i=1}^k p_i, \quad b_k := \sum_{i=1}^k \frac{1}{u_i}, \quad w_k := \frac{1 - q_k}{1 - b_k}$$

とおく. ただし、 $q_0 := 0, b_0 := 0$ とする. さらに、 $1 \leq k \leq n$ を満たす k に対して、

$$r_k := q_k + (1 - b_k)p_k u_k \tag{20}$$

とおく.

定理 3. $p_1 u_1 > 1$ のとき、問題 2 は以下の最適解 (\mathbf{f}^*, w^*) を持つ.

$$f_i^* = \begin{cases} p_i - \frac{w^*}{u_i} & i = 1, \dots, m \\ 0 & i = m + 1, \dots, n \end{cases}$$

$$w^* = w_m$$

ここで、 $m = \max\{k \in \{1, \dots, n-1\} \mid r_k > 1\}$.

Proof. 式 (14), (16) 及び (19) より、 $f_1^* > 0, \dots, f_m^* > 0, f_{m+1}^* = 0, \dots, f_n^* = 0, 1 \leq m < n$ を満たす自然数 m が定まる. まず式 (13) の値が 1 であることを示す. 式 (13) の値を γ とおくと、式 (18) は以下のようになる.

$$\begin{aligned}
\gamma &= \sum_{i=1}^m \frac{p_i}{u_i f_i^* + w^*} + \sum_{i=m+1}^n \frac{p_i}{u_i 0 + w^*} \\
&= \sum_{i=1}^m \frac{\gamma}{u_i} + \sum_{i=m+1}^n \frac{p_i}{w^*} \\
&= \gamma \sum_{i=1}^m \frac{1}{u_i} + \frac{1}{w^*} \sum_{i=m+1}^n p_i \\
&= \gamma b_m + \frac{1}{w^*} (1 - q_m)
\end{aligned}$$

よって $(1 - b_m)\gamma = \frac{1}{w^*}(1 - q_m)$, つまり

$$w^* = \frac{1}{\gamma} \frac{1 - q_m}{1 - b_m} \quad (21)$$

となる. 式 (13) の値を γ とおいたので、 $i = 1, \dots, m$ に対して、

$$\gamma = \frac{p_i u_i}{u_i f_i^* + w^*} = \frac{p_i}{f_i^* + \frac{w^*}{u_i}}.$$

これより

$$f_i^* = \frac{p_i}{\gamma} - \frac{w^*}{u_i} \quad (i = 1, \dots, m) \quad (22)$$

が成立する. 上式を式 (11) へ代入すると、

$$\begin{aligned}
1 &= \sum_{i=1}^m \left(\frac{p_i}{\gamma} - \frac{w^*}{u_i} \right) + w^* \\
&= \frac{1}{\gamma} \sum_{i=1}^m p_i + w^* \left(1 - \sum_{i=1}^m \frac{1}{u_i} \right) \\
&= \frac{1}{\gamma} q_m + w^* (1 - b_m)
\end{aligned}$$

上式に式 (21) を代入すると、

$$1 = \frac{1}{\gamma} q_m + \frac{1}{\gamma} (1 - q_m) = \frac{1}{\gamma}.$$

つまり $\gamma = 1$ が成立する. よって式 (21) より

$$w^* = \frac{1 - q_m}{1 - b_m} = w_m. \quad (23)$$

さらに式 (22) より

$$f_i^* = p_i - \frac{w^*}{u_i} \quad (i = 1, \dots, m).$$

次に、自然数 m の値を特徴付ける. そのために、 $b_m < 1$ かつ $r_{m+1} \leq 1 < r_m$ を示す. 式 (17), (23) 及び $q_m < 1$ より $b_m < 1$ が成立する. 式 (13) の値が 1 なので $i = m$ を代入すれば、

$$\frac{p_m u_m}{u_m f_m^* + w_m} = 1.$$

よって

$$\begin{aligned} p_m u_m &= u_m f_m^* + w_m \\ &> w_m = \frac{1 - q_m}{1 - b_m} \end{aligned}$$

したがって

$$(1 - b_m) p_m u_m > 1 - q_m.$$

つまり

$$q_m + (1 - b_m) p_m u_m > 1.$$

r_k の定義式 (20) より上式は $r_m > 1$ となる. 式 (13) の値が 1 なので、式 (15) へ $i = 1, j = m + 1$ を代入すれば、

$$1 \geq \frac{p_{m+1} u_{m+1}}{w_m}$$

となる. 上式より

$$p_{m+1} u_{m+1} \leq w_m = \frac{1 - q_m}{1 - b_m}.$$

よって

$$(1 - b_m) p_{m+1} u_{m+1} \leq 1 - q_m.$$

つまり

$$q_m + (1 - b_m) p_{m+1} u_{m+1} \leq 1. \quad (24)$$

ここで、 $1 \leq k \leq n$ を満たす k に対して、式 (20) より、

$$\begin{aligned} r_k &= (q_{k-1} + p_k) + \left(1 - \left(b_{k-1} + \frac{1}{u_k} \right) \right) p_k u_k \\ &= q_{k-1} + (1 - b_{k-1}) p_k u_k \end{aligned} \quad (25)$$

上式より式 (24) は $r_{m+1} \leq 1$ となる. 以上より $b_m < 1$ かつ $r_{m+1} \leq 1 < r_m$ が示された.

次に $b_k < 1$ かつ $r_{k+1} \leq 1 < r_k$ を満たす k が存在して一意であることを示す. $b_k < 1$ かつ $1 \leq k < n$ を満たす k に対して、

$$\begin{aligned} r_{k+1} &= q_k + (1 - b_k)p_{k+1}u_{k+1} && \because \text{式 (25) より} \\ &\leq q_k + (1 - b_k)p_k u_k && \because \text{式 (10) より} \\ &= r_k. && \because \text{式 (20) より} \end{aligned}$$

つまり $r_{k+1} \leq r_k$ となり k の一意性が示された. 次に存在を示す. 集合 S を $S = \{k \in \{1, \dots, n-1\} \mid r_k > 1\}$ とおき、 $S \neq \emptyset$ を示す. 式 (20) より

$$\begin{aligned} r_1 &= q_1 + (1 - b_1)p_1 u_1 && \because \text{式 (20) より} \\ &= p_1 + \left(1 - \frac{1}{u_1}\right) p_1 u_1 \\ &= p_1 u_1 > 1. && \because \text{式 (12) より} \end{aligned}$$

よって $r_1 > 1$ なので、 $S \neq \emptyset$ となる. 続いて $k_0 := \max S$ とおき、 $r_{k_0+1} \leq 1$ かつ $b_{k_0} < 1$ を示す. $1 \leq k_0 \leq n-2$ のとき、 $r_{k_0+1} \leq 1$ となる. もしそうでなければ、 k_0 の最大性に矛盾する. $k_0 = n-1$ のとき、

$$\begin{aligned} r_{k_0+1} &= r_n \\ &= q_n + (1 - b_n)p_n u_n && \because \text{式 (20) より} \\ &< 1 && \because \text{式 (9) より} \end{aligned}$$

よって、いずれの場合も $r_{k_0+1} \leq 1$ が成立する. 次に $b_{k_0} < 1$ を示す. $r_{k_0} > 1$ 及び式 (20) より

$$r_{k_0} = q_{k_0} + (1 - b_{k_0})u_{k_0}p_{k_0} > 1.$$

よって

$$(1 - b_{k_0})u_{k_0}p_{k_0} > 1 - q_{k_0} > 0$$

なので、 $1 - b_{k_0} > 0$ つまり $b_{k_0} < 1$ が成立する. 以上より $b_{k_0} < 1$ かつ $r_{k_0+1} \leq 1 < r_{k_0}$ が示された. したがって、一意性により $m = \max\{k \in \{1, \dots, n-1\} \mid r_k > 1\}$ が成立する. \square

次に $p_1 u_1 \leq 1$ の場合を考える.

定理 4. $p_1 u_1 \leq 1$ のとき、問題 2 は最適解 $(f^*, w^*) = (\mathbf{0}, 1)$ をもつ.

Proof. (\mathbf{f}, w) を任意の実行可能解とする. すると

$$\begin{aligned}\mu(\mathbf{f}, w) &= \sum_{i=1}^n p_i \log(u_i f_i + w) \\ &\leq \log \left(\sum_{i=1}^n p_i (u_i f_i + w) \right) \quad \because \log \text{ は上に凸な関数なので Jensen の不等式より} \\ &= \log \left(\sum_{i=1}^n p_i u_i f_i + w \right) \\ &\leq \log \left(\sum_{i=1}^n f_i + w \right) \quad \because p_1 u_1 \leq 1 \text{ 及び式 (10) より} \\ &= \log 1 \quad \because \text{式 (11) より} \\ &= 0.\end{aligned}$$

つまり任意の実行可能解 (\mathbf{f}, w) に対して、 $\mu(\mathbf{f}, w) \leq 0$ が成立する. $\mu(\mathbf{0}, 1) = 0$ なので、問題 2 は $(\mathbf{f}^*, w^*) = (\mathbf{0}, 1)$ において最適解を持つ. \square

5 機械学習

近年では機械学習を用いた研究が行われています. 私のお気に入りの論文は、Benter (2008), Edelman (2007), Lessmann et al. (2009), Lessmann et al. (2010), Lessmann et al. (2012) です. conditional logit モデルの基本について知りたい方は、Train (2009) の第 3 章を参考にすると良いです.

参考文献

- Ali, M. M. (2008). Probability and utility estimates for racetrack bettors. In *Efficiency Of Racetrack Betting Markets*, pages 71–83. World Scientific.
- Benter, W. (2008). Computer based horse race handicapping and wagering systems: a report. In *Efficiency of racetrack betting markets*, pages 183–198. World Scientific.
- Edelman, D. (2007). Adapting support vector machine methods for horserace odds prediction. *Annals of Operations Research*, 151(1):325–336.
- Ethier, S. (2010). *The Doctrine of Chances: Probabilistic Aspects of Gambling*. Probability and Its Applications. Springer Berlin Heidelberg.
- Körner, T. (2008). *Naive Decision Making: Mathematics Applied to the Social World*. Cambridge University Press.
- Lessmann, S., Sung, M.-C., and Johnson, J. E. (2009). Identifying winners of competitive events: A svm-based classification model for horserace prediction. *European Journal of Operational Research*, 196(2):569–577.
- Lessmann, S., Sung, M.-C., and Johnson, J. E. (2010). Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting*, 26(3):518–536.
- Lessmann, S., Sung, M.-C., Johnson, J. E., and Ma, T. (2012). A new methodology for generating and combining statistical forecasting models to enhance competitive event prediction. *European Journal of Operational Research*, 218(1):163–174.
- Lo, V. S. and Bacon-Shone, J. (2008a). Approximating the ordering probabilities of multi-entry competitions by a simple method. In *Handbook of Sports and Lottery Markets*, pages 51–65. Elsevier.
- Lo, V. S. and Bacon-Shone, J. (2008b). Probability and statistical models for racing. *Journal of Quantitative Analysis in Sports*, 4(2).
- Tijms, H. (2012). *Understanding Probability*. Cambridge University Press.
- Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.